Structure-based design of ligands for vitamin transporters in bacteria

Monjas Gómez, Leticia

Chapter 5

Dynamic combinatorial chemistry to identify binders of ThiT

Dynamic combinatorial chemistry (DCC) is a powerful method for the identification of ligands for (protein) targets. In this chapter, we applied DCC to find binders of ThiT. To do so, we designed a small library of acylhydrazones and we explored different methods for the analysis of the library. Our attempts to isolate the protein–binder complexes by using techniques such as affinity chromatography or size-exclusion chromatography failed. When we used saturation-transfer difference (STD) NMR spectroscopy, we were able to identify six binders of ThiT from the library. Preliminary data indicate that the acylhydrazones bind to ThiT with K_D values in the micromolar range.

The introduction of this chapter is adapted from the original publication:
Monjas, L.; Hirsch, A. K. H. *Future Med. Chem.* **2015**, *7*, 2095.

Part of the results of this chapter will be submitted for publication:
Monjas, L.; Swier, L. J. Y. M.; Slotboom, D. J.; Hirsch, A. K. H. *Manuscript in preparation.*

5.1 Introduction

Dynamic combinatorial chemistry (DCC) is a tool to accelerate the drug-discovery process and it holds the potential to accelerate both hit identification and optimization. DCC was first introduced in the 1990s,[1,2] and since then, numerous examples of applications in medicinal chemistry and chemical biology have been reported.[3–6]

DCC is a method in which different building blocks react reversibly to generate libraries of chemical compounds under thermodynamic control (Figure 1). A dynamic combinatorial library (DCL) is adaptive if it is formed in the presence of the target, which works as a template, or pre-equilibrated, if the target is added once the DCL has reached equilibrium. The former is preferred, given that it allows taking maximum advantage of the DCC methodology. Here, the system can adapt due to molecular recognition events between the protein target and one or more of the library members, the equilibrium will shift, resulting in amplification of the best binders at the expense of other members of the DCL that contain the same building blocks. Pre-equilibrated DCLs are especially useful when working with targets that are not compatible with the equilibration conditions.

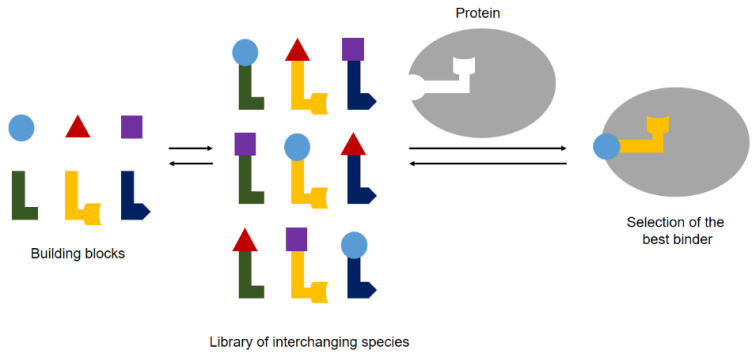


Figure 1. Schematic representation of protein-templated dynamic combinatorial chemistry (DCC).

To generate a DCL, it is possible to use a range of chemical reactions; however, some criteria must be fulfilled. The reaction of choice should be biocompatible, meaning that it should take place in aqueous medium and at a temperature and pH at which the protein target is stable (for adaptive DCLs). It is also important that the reactions are chemoselective to avoid the formation of undesired compounds or side reactions with the protein. The building blocks should have similar reactivity and energy to ensure that comparable amounts of each library member are formed in an untemplated DCL. Furthermore, all members of the DCL must be soluble, to avoid that any compound would

precipitate, biasing the equilibrium. To improve the solubility, a small percentage of a cosolvent, such as DMSO, can be used if tolerated by the protein target. In some instances, when the reaction of choice is slow, the use of a catalyst can speed up the reaction to ensure it proceeds in a time frame during which the protein remains folded. For example, a nucleophilic catalyst such as aniline can be used when the reversible system consists of acylhydrazones, promoting the equilibration of the DCL at a pH compatible with most protein targets rather than the acidic pH required in the absence of a catalyst.[7,8]

To analyze the library, the conditions should be modified in order to 'freeze' the equilibrium and ensure that the composition remains constant. This can be done by changing the pH or irreversible modification of the ligands (for example, *in situ* reduction of imines to amines). For relatively small libraries, separation techniques such as HPLC, HPCE (high-performance capillary electrophoresis) or GC are routinely applied. In some cases, MS- or NMR-based techniques have been successfully used. For larger libraries, a combination of chromatographic techniques coupled to MS can simplify the analysis.[9]

In most reports of protein-templated DCC, the potentially bioactive ligands feature reversible covalent bonds connecting the various building blocks, and these ligands interact with the biological target through non-covalent bonds. Once a hit has been identified, bioisosteric modification of the reversible connector is necessary in most cases to obtain stable analogues whilst preserving the biochemical activity.

The increasing number of publications over the past twenty years is a good reflection of the success of DCC applied to medicinal chemistry and chemical biology. DCC has enabled the discovery of binders of a variety of DNA, RNA and protein targets, including several enzyme classes, such as proteases, anhydrases, kinases and oxygenases.[5]

The main advantage of DCC is its potential to accelerate the hit-identification process: library generation and screening of the potential ligands are combined into a single operation, avoiding the individual synthesis, purification, characterization and biochemical evaluation of every single member of the DCL.

The most challenging part of DCC is often the analysis, which becomes more difficult with increasing size and complexity of the DCL. Up to now, mainly small libraries and well-known targets have been studied. Long equilibration times are another limitation, especially for unstable protein targets or library members.

New approaches have emerged to circumvent the limitations of DCC. For example, the analysis of large and more complex libraries can be simplified by looking only at the bound species by using mild techniques to identify protein–ligand complexes or by 'fishing' these complexes from the mixture. For

instance, size-exclusion chromatography (SEC) has been used for the separation of the ligand–target complexes from the DCL. After denaturation, the ligands are released from the protein and identified by MS analysis.[10–12]

Other options are the immobilization of the target or library members on a solid support. For example, binders of the carbohydrate-binding protein concanavalin A have been identified by adding the immobilized protein target to the DCL.[13] The protein was immobilized on sepharose beads, allowing for the removal of the non-binding compounds by filtration, and subsequent elution of the bound compounds for characterization. The reverse approach, immobilization of the library members, has been also reported. In this so-called resin-bound DCC (RBDCC),[14] the building blocks are attached to a resin and combined with the same building blocks in solution to form a DCL of dimers. After incubation with the fluorescently-tagged DNA target, wash and analysis of the fluorescent beads, DNA binders were successfully identified. There are a few more proof-of-principle studies regarding the immobilization of the target or library components on a solid support, but these are also carried out using model systems and do not involve protein targets.[15–17] Alternatively, by choosing the right conditions for imine formation[18,19] or a reversible reaction that leads to inherently labile products such as hemithioacetals,[20] virtual dynamic combinatorial libraries can be generated. In these systems, little or no products are formed in the absence of the target protein, meaning that only binders have to be detected, thereby facilitating the analytical challenge.

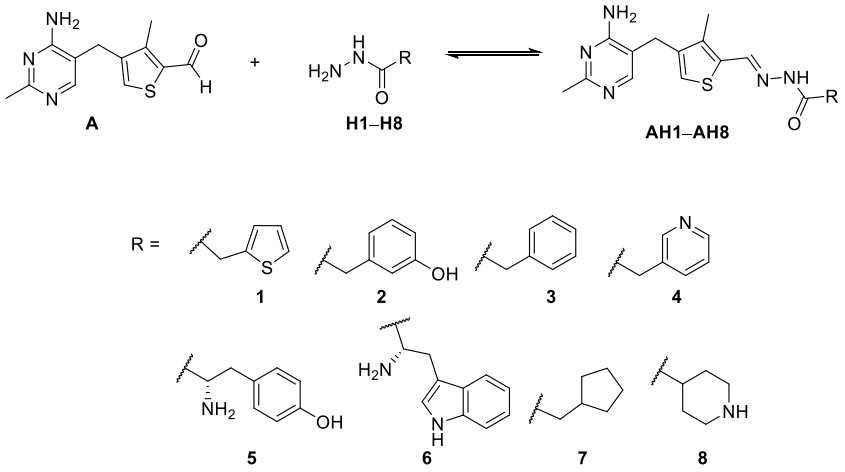
There are a few reports in which challenging targets have been studied. For example, competitive MS-binding assays have been developed to identify binders of the γ-aminobutyric acid (GABA) transporter 1 (GAT1).[21] This method is highly sensitive, allowing the detection of binders using a low concentration of the target protein. This approach can be especially useful when studying targets such as G-protein coupled receptors, ion channels and transporters. It requires, however, a native MS marker for the protein of interest, and deconvolution experiments after library screening, therefore it is not suitable for big DCLs.

Despite the progress, there are still a number of challenges that need to be addressed. The analysis of the DCLs is one of the most important bottlenecks. In this chapter, efforts towards new methods for the isolation of the protein–ligand complexes are described, as well as the use of saturation-transfer difference (STD) NMR spectroscopy for the identification of binders of ThiT.

5.2 Results and discussion

5.2.1 Design of the dynamic combinatorial library: acylhydrazones as binders of ThiT

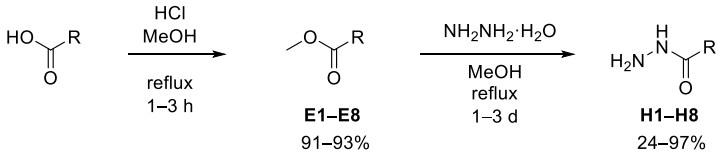
In order to explore the binding pocket of ThiT at the hydroxyl end of thiamine (as in Chapter 4), we designed compounds featuring a thiophenyl instead of a thiazolium ring, as in previous chapters, and introduced an acylhydrazone motif as a linker to enable DCC. We used the software MOLOC[22] and LeadIT[23] for modeling and docking, respectively, and the scoring function HYDE,[24,25] as described in Chapter 2, Section 2.2.1. We used the crystal structure of ThiT in complex with thiamine (PDB ID: 3RLB).[26] First, we modeled aromatic moieties linked directly to the acylhydrazone functionality, given that there are two aromatic residues (Trp63 and Tyr85) with which the new compound could interact *via* π–π-stacking interactions. All the attempts were, however, unsuccessful due to high intermolecular clashes of the designed compounds and ThiT, probably due to the rigidity of the molecules, which were 'too flat' because of the aromatic moiety being in conjugation with the acylhydrazone linker. The problem was solved by introducing a methylene linker: most of the modeled molecules were accommodated in the pocket and were engaged in π–π-stacking interactions with Trp63 and Tyr85 in most of the cases. We also modeled aliphatic substituents, and from all the docked molecules, we selected eight acylhydrazones (**AH1**–**AH8**) with ΔG_{est} = –50 to –56 kJ mol^{-1} for the *E* isomer, which can be obtained by reaction of aldehyde **A** and hydrazides **H1**–**H8** (Scheme 1).



Scheme 1. Designed dynamic combinatorial library (DCL) to afford acylhydrazones as binders of ThiT.

5.2.2 Synthesis of the building blocks

We synthesized aldehyde **A** as described in Chapter 2, Section 2.2.2. Hydrazides **H1–H8** were obtained from their corresponding methyl esters (ethyl ester in the case of **H6**), which were commercially available or synthesized by esterification of the corresponding carboxylic acid, using HCl in methanol at reflux, in 91–93% yield (Scheme 2). Next, the reaction of the corresponding ester with hydrazine monohydrate at reflux afforded the hydrazides in 24–97% yield. For seven of the hydrazides, the yield of this reaction was high (82–97%); only for **H8** the yield was low (24%) because of problems during the recrystallization.



Scheme 2. Synthesis of the hydrazides **H1–H8**.

5.2.3 Formation and analysis of the DCL to identify ThiT binders

We explored three different methods in order to analyze the DCL for the identification of ThiT binders, although some unsuccessfully: affinity chromatography (nickel-sepharose column), SEC and STD-NMR spectroscopy.

5.2.3.1 Affinity chromatography: nickel-sepharose column

The initial idea was to isolate the protein–ligand complexes from the DCL, to facilitate the identification of the binders. In this way, the analytical challenge is simplified: because the non-binders are removed from the system, only the binders need to be identified. We had planned to add ThiT with an *N*-terminal His_8-tag to the preformed DCL consisting of acylhydrazones, and let the mixture incubate, enabling the formed acylhydrazones to bind to ThiT. To ensure that the side chains of the His residues (pK_a 6.0) of the His_8-tag are not protonated, leading to the intended interaction of the His_8-tag of ThiT with nickel, the pH was increased from 5.0 (at which the DCL is formed) to 7.0. In this case, we would have a pre-equilibrated DCL, given that the equilibrium is 'frozen' upon raising the pH. As mentioned in Section 5.1, an adaptive DCL has the advantage of adjusting its composition after addition of the protein, resulting in amplification of the best binder(s). Given that our idea is to isolate the protein–ligand complexes, this is not required as we do not need to determine amplification factors. Subsequently, the mixture of ThiT and DCL

would be incubated with nickel-sepharose beads, leading to immobilization of ThiT through coordination of the nickel by the His_8-tag. Next, a buffer containing a low concentration of imidazole would be used to elute the non-binders from the nickel-sepharose column, followed by a buffer with a high concentration of imidazole to elute ThiT in complex with the binders (Figure 2). Imidazole is used in the buffers because it can compete with the His_8-tag for binding to the nickel-sepharose beads.

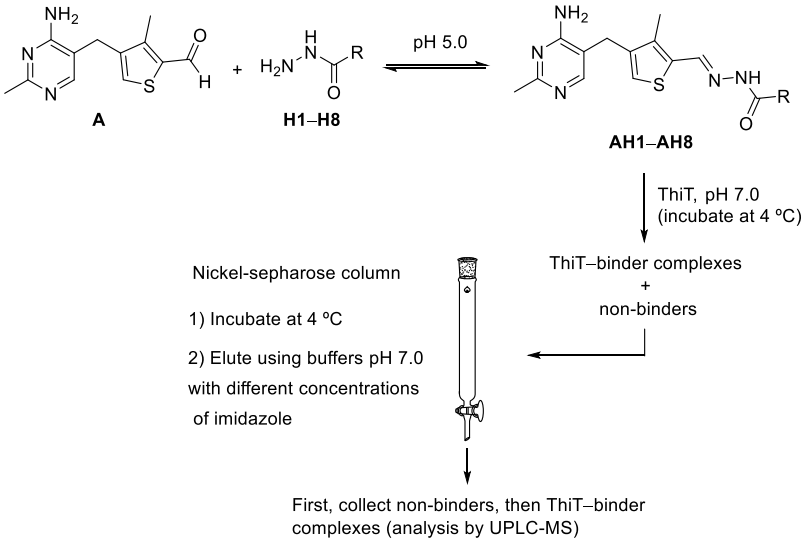


Figure 2. Planned strategy to identify binders of ThiT using a nickel-sepharose column.

First, we did a control experiment without protein, following the procedure depicted in Figure 2. We prepared the DCL by mixing the aldehyde **A** and the eight hydrazides **H1–H8** at pH 5.0 and incubating the mixture at room temperature for 24 h, using a rotary mixer. Subsequently, upon raising the pH of the DCL, we incubated it with nickel-sepharose beads. Then, we used two buffers with increasing concentration of imidazole, 50 mM and 500 mM, which are the standard conditions used to purify ThiT. We analyzed the collected fractions by UPLC-MS, and we found acylhydrazones in all the fractions. These results indicate that these concentrations of imidazole are not suitable for the experiment with protein, given that the non-binders would also elute with the ThiT–binder complexes.

In order to find conditions for the nickel-sepharose column, we did a new control experiment, following the exact same procedure, but using buffers with varying concentrations of imidazole for the elution of the nickel-sepharose column: we used ten buffers of increasing concentration of imidazole: 50, 100, 150, 200…500 mM imidazole. After analyzing the samples by UPLC-MS, we

observed that the last fraction containing acylhydrazones was the one eluted using 250 mM imidazole.

The next control experiment was to check whether ThiT was still bound to the nickel-sepharose column using buffer containing 250 mM imidazole. We repeated the experiment now with ThiT and without the library. After eluting the nickel-sepharose column with buffer containing 250 mM imidazole, about one third of the amount of ThiT eluted out of the column. Therefore we could not achieve good separation of the non-binders from the ThiT-binder complexes.

Given these unsuccessful results, we decided to modify our strategy and use a different method to separate the non-binders from the ThiT–binder complexes.

5.2.3.2 Size-exclusion chromatography (SEC)

As mentioned in Section 5.1, SEC in combination with MS analysis has been used to identify protein ligands from a DCL.[10-12] We thought that we could use a similar approach for our target protein. We planned the experiment in the same way as for the nickel-sepharose column, but using SEC instead to isolate the ThiT–binder complexes (Figure 3). In this case, the order of elution would be the opposite as for the nickel-sepharose column: the ThiT–binder complexes should elute first followed by the non-binders.

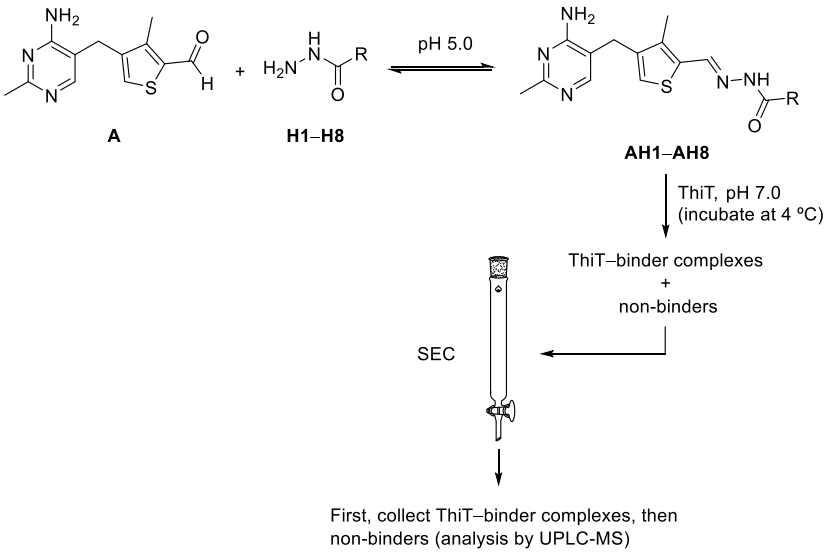


Figure 3. Planned strategy to identify binders of ThiT using size-exclusion chromatography (SEC).

We prepared the DCL by mixing the aldehyde **A** and the eight hydrazides **H1–H8** at pH 5.0, and incubating the mixture at room temperature for 24 h in a rotary mixer. Then, we divided the DCL into two parts: for a control experiment and for the experiment with ThiT. For the experiment with ThiT, we incubated the DCL and ThiT at 4 °C for 2 h. Then, we centrifuged the vial, and some aggregates appeared. In the control experiment, there were no aggregates after centrifugation, so we assumed that the aggregates corresponds to ThiT or ThiT in complex with some of the binders. We did the SEC of the samples and achieved perfect separation of ThiT from the compounds of the library (Figure 4).

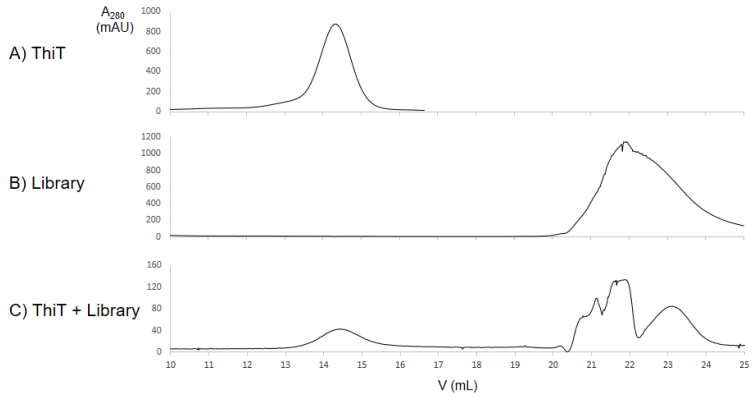


Figure 4. UV profiles (λ = 280 nm) of the collected fractions of the SEC. **A)** Purification of ThiT. **B)** Control experiment (only the library). **C)** Experiment using ThiT and the library.

We analyzed the fractions of the control experiment and of the experiment with ThiT (both peaks and also a solution of the aggregates that were formed after incubation) by UPLC-MS. To benefit from better sensitivity enabling the detection of very small amounts of compounds, we used a selective-ion-recording (SIR) method (UPLC-TQD-SIR). As expected, we could identify the acylhydrazones in the control experiment (Figure 4B), in the experiment with ThiT (Figure 4C, peak 20–25 mL) as well as in the aggregates. However, in the peak in which ThiT elutes in the experiment (Figure 4C, peak 13–16 mL), after adding acetonitrile to denature the protein and release the binders, we could not detect any acylhydrazone. Presumably, the amount of protein was too low (initial concentration of 10 μM) to be able to identify compounds binding to it.

Given that SEC did not enable identification of binders of ThiT, we decided to change the approach, and use an analytical technique requiring a lower amount of protein.

5.2.3.3 ^{1}H-STD-NMR spectroscopy

STD-NMR is a powerful technique to study protein–ligand interactions in solution (Figure 5). When a protein is saturated by selective irradiation, the ligand(s) that bind to this protein become also saturated by spin diffusion, given that the bound ligand(s) behave as part of the protein, with reduced mobility in solution. By chemical exchange, that saturation can be detected when the ligand is again free in solution. When subtracting the spectrum measured while saturating the protein (*on*-resonance) from the reference spectrum (*off*-resonance), the difference spectrum is obtained (STD), which only features the signals of the ligand(s) that bind to the protein (Figure 5). The non-binders do not receive saturation from the protein, therefore their signals are exactly the same as in the reference spectrum and they do not appear in the STD-NMR spectrum. The protein has to be irradiated at a frequency where only the protein resonances, and not the ligands. Usually, the concentration of ligand(s) is 10–100-fold the amount of protein, which allows to work with very low concentration of protein (in the micromolar range).[27]

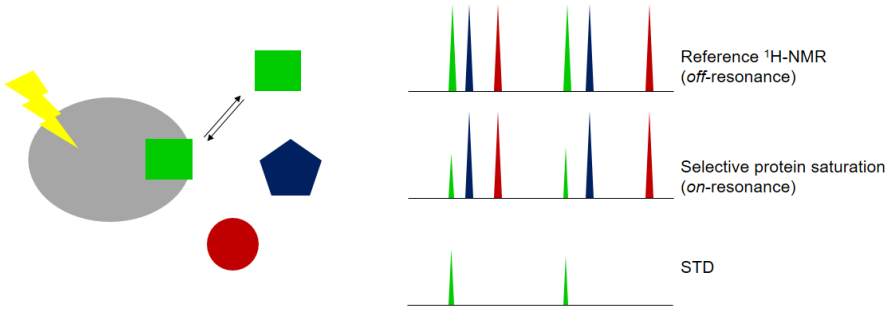


Figure 5. Principle of saturation-transfer difference (STD)-NMR to study protein–ligand interactions. Color code: protein: gray, binder: green, non-binders: blue and red.

^{1}H-STD-NMR spectroscopy has been successfully applied in combination with DCC in some studies,[20,28] and in a few cases to study ligand binding to transmembrane proteins.[29,30] For our target, we did first a control experiment with ThiT and a known binder (**B1**) from Chapter 2 (Figure 6). This control experiment is necessary for two reasons: first, the available amount of protein is not enough to record a ^{1}H-NMR spectrum to identify a suitable irradiation frequency; second, we wanted to establish whether the conditions are optimal and enable us to detect a known binder, given that in previous reports in which STD-NMR spectroscopy has been applied to transmembrane proteins, they were embedded into the lipid bilayer of a liposome, but not in solution with detergents as in our case. To obtain a sample of ThiT in

deuterated buffer (pD 7.0), we performed a buffer-exchange column, providing us with 400 µL of 9.8 µM ThiT, which corresponds to a final concentration of ThiT in the NMR tube of 7.8 µM for 500 µL. Running the ^{1}H-STD-NMR experiment using a 20-fold excess of ligand did not lead to any clearly detectable signal, upon irradiation either at –1 or –2 ppm. Using a 100-fold excess of ligand, and measuring for 11 h irradiating at –1 ppm resulted in a difference spectrum featuring peaks corresponding to our known binder **B1**, as well as the detergent (*n*-decyl-β-D-maltopyranoside) that is present in the buffer (Figure 6). We obtained the same ^{1}H-STD-NMR spectrum when irradiating at –2 ppm, so we could choose any of these frequencies for the experiment with the DCL.

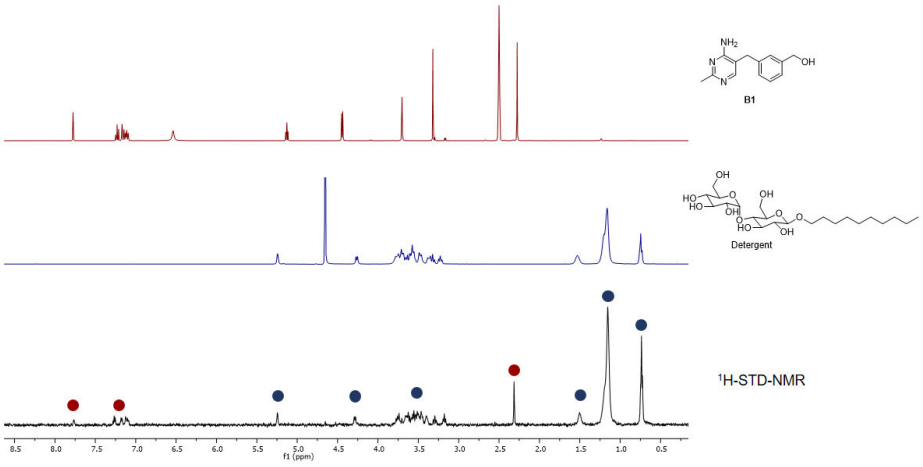


Figure 6. Control experiment. The ^{1}H-NMR spectra are: in brown binder **B1** in DMSO-d_6, in blue the detergent in D_2O and in black the ^{1}H-STD-NMR of the control experiment (buffer in D_2O with 5% DMSO-d_6), in which peaks corresponding to **B1** or detergent are labeled with a circle according to the color of their spectrum.

We decided to perform the experiment with our library of compounds using these conditions (100-fold excess of ligand, irradiation at –1.1 ppm, 11 h). To analyze our DCL, we divided it into two sublibraries, containing the aldehyde and four hydrazides each, selected in a way that the characteristic peaks do not overlap. First, we prepared samples with individual acylhydrazones **AH1–AH8** by reaction of aldehyde **A** and the corresponding hydrazide **H1–H8** at pD 5.0, using an excess of aldehyde **A** (5.0 eq) to ensure that the hydrazide is consumed and the aromatic peaks can be assigned to the acylhydrazone, given that the aldehyde only has two peaks in the aromatic region (7.7 and 7.6 ppm).

For the experiment with ThiT, the building blocks were left to react in a buffer at pD 5.0 for 24 h, in a rotary mixer. Then, the DCL was added to the

solution of ThiT in a buffer at pD 7.0, and the *on*-resonance and *off*-resonance spectra were recorded. For the first library, DCL-A, we included the aldehyde **A** and the hydrazides **H1**, **H2**, **H6** and **H7**. The second library, DCL-B, consists of the same aldehyde **A** and the hydrazides **H3**, **H4**, **H5** and **H8** (Figure 7).

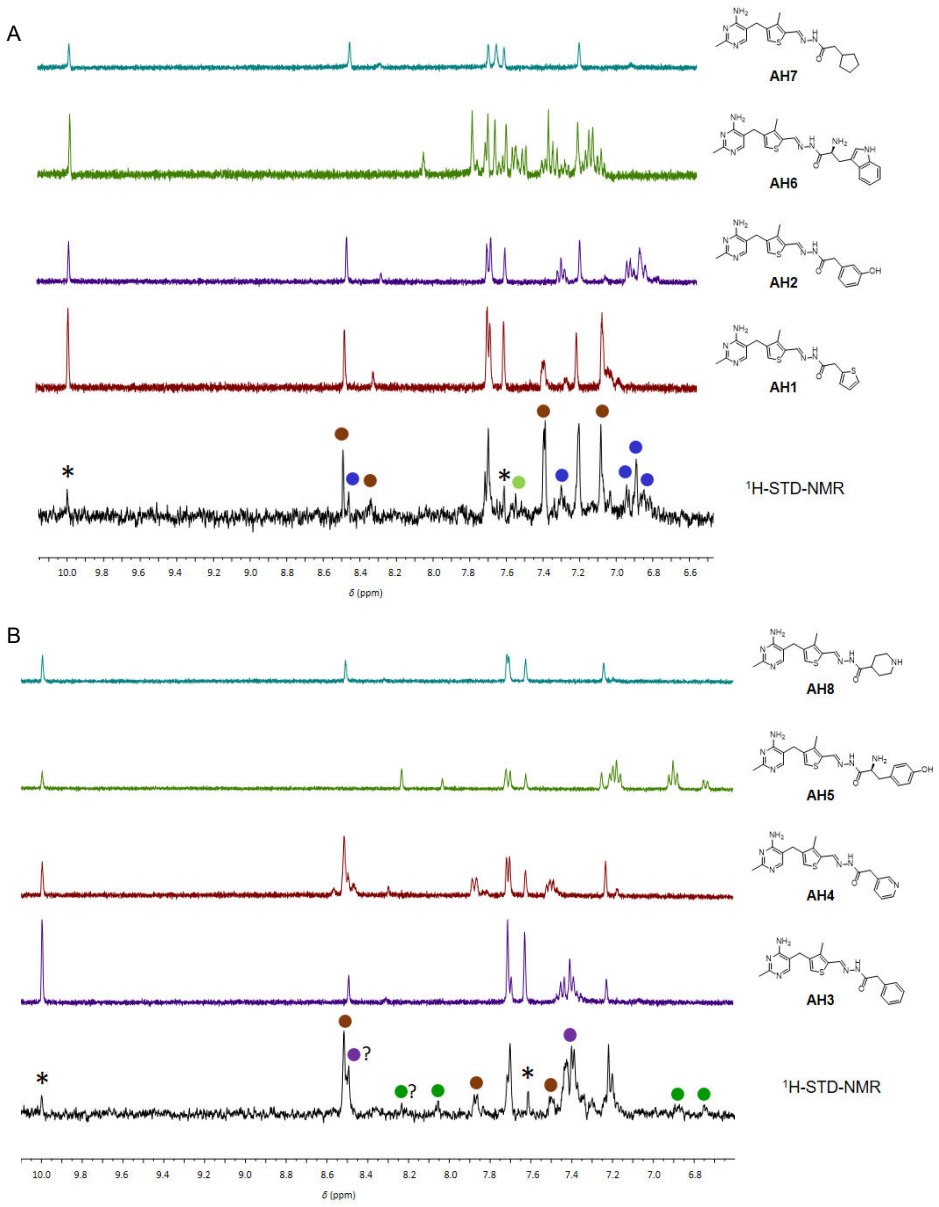
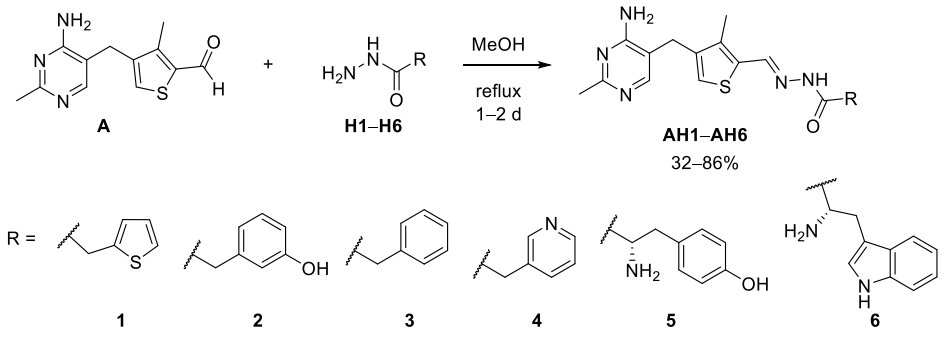


Figure 7. A) ¹H-NMR spectra of four individual acylhydrazones (obtained using an excess of aldehyde *vs* hydrazide), used in the DCL-A. At the bottom, ¹H-STD-NMR spectrum of DCL-A, in which peaks corresponding to the aldehyde and each acylhydrazone are labeled with a star and a circle according to the color of their spectrum. **B)** Same as for A, but for DCL-B.

In DCL-A, we observed that three of the four possible acylhydrazones bind to ThiT (Figure 7A). In the ^{1}H-STD-NMR spectrum we could see peaks at 10.0, 7.7 and 7.6 ppm corresponding to the aldehyde **A**, and the peaks corresponding to the protons of the pyrimidinyl and the thiophenyl rings at ~7.7 and ~7.2 ppm, common to the four acylhydrazones. In principle, after 24 h of incubation to form the DCL, all the aldehyde **A** was consumed according to ^{1}H-NMR, but apparently, a small amount was still in solution, and after many scans during the experiment we could see it in the ^{1}H-STD-NMR spectrum. We could identify compound **AH1** by its characteristic peaks at ~7.4 and ~7.1 ppm and **AH2** especially by its signals at ~7.3 and 7.0–6.8 ppm. Identification of **AH6** was a bit more challenging, but the peaks at ~7.5 ppm in the ^{1}H-STD-NMR spectrum could be only assigned to this compound. Compound **AH7** does not have characteristic peaks in the aromatic region, but when we analyzed the aliphatic part, we did not observe any of its peaks in the ^{1}H-STD-NMR spectrum. In DCL-B, three of the other four possible acylhydrazones turned out to bind to ThiT (Figure 7B). The peaks at ~7.4 ppm, and probably the one at ~8.5 ppm in the ^{1}H-STD-NMR spectrum indicate that compound **AH3** binds to ThiT. Compound **AH4** was identified by its signals at ~8.50, ~7.85 and ~7.50 ppm. Compound **AH5** was easily identified in the ^{1}H-STD-NMR spectrum by its signals at ~6.7 and ~6.9 ppm, and also because of the peaks at 8.05 and probably at 8.20 ppm. Compound **AH8** has no characteristic peaks in the aromatic region, and when we analyzed the aliphatic part we did not observe any peak in the ^{1}H-STD-NMR spectrum. Taken together, the analysis of both libraries shows that all the aromatic derivatives (**AH1–AH6**) bind to ThiT, whereas the aliphatic ones (**AH7** and **AH8**) do not bind.

5.2.4 Synthesis of the identified binders

We synthesized the six identified binders by reaction of aldehyde **A** and the corresponding hydrazide **H1–H6** at reflux in MeOH, affording the desired products **AH1–AH6** in 32–86% yield (Scheme 3).



Scheme 3. Synthesis of the acylhydrazones identified as binders of ThiT.

5.2.5 Binding-affinity determination of the identified binders

We determined the binding affinity of four of the six acylhydrazones identified as binders of ThiT by isothermal titration calorimetry (ITC), given that the compounds are fluorescent and we could not use the intrinsic-protein-fluorescence titration assay as described in Chapter 2. We tested the compounds as mixtures of *E*/*Z* isomers. The preliminary results are shown in Table 1. We observed that acylhydrazones **AH1** and **AH5** bind to ThiT with K_D values in the low-micromolar range, while **AH3** and **AH4** are slightly weaker binders. The binding affinities are weaker than predicted, although we have to consider that we tested the acylhydrazones as mixtures of *E* and *Z* isomers, and only the *E* isomer is predicted to be a strong binder of ThiT.

Table 1. The binding affinities of ThiT for the acylhydrazones, together with the experimental (ΔG_{exp}) and estimated Gibbs free energies of binding (ΔG_{est}). The estimated values are based on the scoring function HYDE for the *E* isomer.

Compound	K_D (μM)	ΔG_{exp} (kJ mol^{-1})	ΔG_{est} (kJ mol^{-1})
AH1	3.85[a]	–31	–51
AH2	n.d.	n.d.	–56
AH3	30.2[a]	–26	–53
AH4	22.0[a]	–27	–52
AH5	3.68[a]	–31	–53
AH6	n.d.	n.d.	–53

n.d. = not determined

[a] The K_D value was determined from 1 experiment.

5.3 Conclusions

We have shown that it is necessary to have a robust protein to develop a new method for the isolation of protein–ligand complexes from a DCL, as well as a relatively large amount of protein, because it is necessary to perform a lot of control experiments before doing the 'real' DCC experiment. The amount of ThiT that we had available did not allow us to identify the right conditions to establish two of these methodologies. In the case of the nickel-sepharose column, it was not possible to separate the fraction containing the non-binders from the fraction of ThiT–binder complexes. Larger amounts of protein would have enabled us to perform more control experiments by modifying the amount of stationary phase, and then finding the optimal concentration of imidazole for the wash and elution buffers. In the case of SEC, we achieved perfect baseline separation of the ThiT–binder complexes from the non-binders. Yet, we faced detection problems by UPLC-MS, even using the highly sensitive SIR method.

For ^{1}H-STD-NMR spectroscopy, the amount of protein required for an experiment is smaller than for the other two techniques, enabling us to identify binders of ThiT from a DCL. The disadvantages of this method are the limited size of the DCL and that it is necessary to determine the ^{1}H-NMR reference spectrum of each individual product for comparison with the ^{1}H-STD-NMR spectrum. The advantages are that it is a very simple technique and that the amount of protein necessary to succeed is very small. Preliminary results indicate that the acylhydrazones identified by ^{1}H-STD-NMR bind to ThiT with K_D values in the micromolar range. If we compare their binding affinities with compounds in Chapter 4, in which most of the compounds bind in the nanomolar range, we could conclude that the acylhydrazones are weaker binders. We need to take into account, however, that the K_D values were determined for mixtures of *E*/*Z* isomers, and only the *E* isomer is predicted to bind with high affinity to ThiT.

5.4 Experimental section

5.4.1 Modeling and docking

For general experimental details, see Chapter 2, Section 2.4.1. The only difference is that for the compounds of this chapter, during docking, the binding site in the protein was restricted to 15.0 Å around the cocrystallized thiamine.

5.4.2 Synthesis

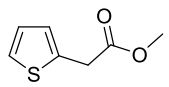
General methods. For general experimental details, see Chapter 2, Section 2.4.2. Optical rotations were measured on a Schmidt & Haensch polarimeter (Polartronic MH8) with a 10 cm cell (*c* given in g/100 mL). Aldehyde **A** was synthesized as described in Chapter 2, Section 2.4.2 (compound **9**). Esters **E3**, **E5**, **E6** and **E8** are commercially available.

General procedure for the synthesis of esters E1, E2, E4 and E7 (GP-A). To a solution of the corresponding carboxylic acid (1.0 eq) in anhydrous MeOH (0.1 M), HCl·MeOH (1.25 M HCl in MeOH, 0.2 eq) was added, and the reaction mixture was stirred at reflux (90 °C, pre-heated oil bath) for 1–3 h. Then, a saturated aqueous solution of $NaHCO_3$ was added, and the reaction mixture was extracted 3 times with CH_2Cl_2, the combined organic layers were washed once with water, once with a saturated aqueous solution of NaCl, dried over $MgSO_4$, filtered, and concentrated under reduced pressure. The corresponding esters were obtained in 91–93% yield.

General procedure for the synthesis of hydrazides H1–H8 (GP-B). To a solution of the corresponding ester (1.0 eq) in MeOH or EtOH (0.1 M, EtOH only in the case of **H6**, for which the starting material was the commercially available ethyl ester), $NH_2NH_2 \cdot H_2O$ (NH_2NH_2 64–65%, 2.0–4.0 eq) was added, and the reaction mixture was stirred at reflux (90 °C for MeOH or 105 °C for EtOH, pre-heated oil bath) for 1–3 days. Then, the reaction mixture was concentrated under reduced pressure, and the crude was purified as indicated in each case. The corresponding hydrazides were obtained in 24–97% yield. Hydrazide **H6** was synthesized according to **GP-B** and its spectroscopic data correspond to those reported in the literature.[31]

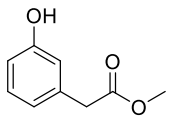
General procedure for the synthesis of acylhydrazones AH1–AH6 (GP-C). To a solution of aldehyde **A** (1.0 eq) in MeOH (*ca.* 0.07 M), the corresponding hydrazide (1.2 eq; except for compound **AH4**, see specific procedure) was added, and the reaction mixture was stirred at reflux (90 °C, pre-heated oil bath) for 1–2 days. Then, the reaction mixture was concentrated under reduced pressure, and the crude was purified by flash column chromatography. The corresponding acylhydrazones were obtained as mixtures of *E*/*Z* isomers in 32–86% yield, and the peaks of both isomers are reported in the ^{1}H- and ^{13}C-NMR spectra.

Methyl 2-(thiophen-2-yl)acetate (E1):



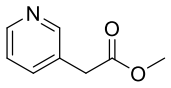
This compound was synthesized according to **GP-A**, starting with 2-(thiophen-2-yl)acetic acid (0.213 g, 1.50 mmol). Ester **E1** was obtained as a yellow oil (0.219 g, 1.40 mmol, 93%). ^{1}H-NMR (400 MHz, $CDCl_3$) δ 7.22 (dd, J = 4.9, 1.5, 1H), 6.97–6.94 (m, 2H), 3.85 (s, 2H), 3.73 (s, 3H). ^{13}C-NMR (101 MHz, $CDCl_3$) δ 171.0 (C), 135.1 (C), 126.95 (CH), 126.93 (CH), 125.2 (CH), 52.4 (CH_3), 35.3 (CH_2). IR (cm^{-1}) 3075, 3004, 2953, 1736, 1436, 1330, 1235, 1167, 1008, 852, 696. HRMS (ESI–) calculated for $C_7H_7O_2S$ [M – H]$^-$ 155.0161, found 155.0174.

Methyl 2-(3-hydroxyphenyl)acetate (E2):



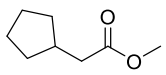
This compound was synthesized according to **GP-A**, starting with 2-(3-hydroxyphenyl)acetic acid (0.228 g, 1.50 mmol). Ester **E2** was obtained as a yellow oil (0.228 g, 1.37 mmol, 91%). ^{1}H-NMR (400 MHz, $CDCl_3$) δ 7.19 (t, J = 7.8, 1H), 6.85–6.83 (m, 1H), 6.78–6.73 (m, 2H), 3.70 (s, 3H), 3.58 (s, 2H). ^{13}C-NMR (101 MHz, $CDCl_3$) δ 172.9 (C), 156.2 (C), 135.3 (C), 129.9 (CH), 121.5 (CH), 116.4 (CH), 114.5 (CH), 52.4 (CH_3), 41.2 (CH_2). IR (cm^{-1}) 3396, 3027, 2955, 1712, 1589, 1455, 1437, 1278, 1215, 1151, 1000, 960, 762, 689. HRMS (ESI+) calculated for $C_9H_{11}O_3$ [M + H]$^+$ 167.0703, found 167.0702.

Methyl 2-(pyridin-3-yl)acetate (E4):



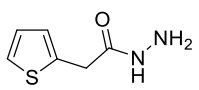
This compound was synthesized according to **GP-A**, starting with 2-(pyridin-3-yl)acetic acid (1.32 g, 7.60 mmol). Ester **E4** was obtained as a yellow oil (1.06 g, 7.01 mmol, 92%). ^{1}H-NMR (400 MHz, $CDCl_3$) δ 8.54 (s, 2H), 7.71–7.68 (m, 1H), 7.33–7.29 (m, 1H), 3.72 (s, 3H), 3.66 (s, 2H). ^{13}C-NMR (101 MHz, $CDCl_3$) δ 171.1 (C), 150.3 (CH), 148.5 (CH), 136.9 (CH), 129.8 (C), 123.5 (CH), 52.3 (CH_3), 38.3 (CH_2). IR (cm^{-1}) 3033, 3004, 2954, 1733, 1577, 1481, 1426, 1234, 1161, 797, 711. HRMS (ESI+) calculated for $C_8H_{10}NO_2$ [*M* + H]$^+$ 152.0706, found 152.0706.

Methyl 2-cyclopentylacetate (E7):



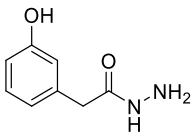
This compound was synthesized according to **GP-A**, starting with 2-cyclopentylacetic acid (0.128 g, 1.00 mmol). Ester **E7** was obtained as a colorless oil (0.132 g, 0.928 mmol, 92%). ^{1}H-NMR (400 MHz, $CDCl_3$) δ 3.62 (s, 3H), 2.28–2.26 (m, 2H), 2.24–2.12 (m, 1H), 1.82–1.74 (m, 2H), 1.63–1.45 (m, 4H), 1.15–1.06 (m, 2H). ^{13}C-NMR (101 MHz, $CDCl_3$) δ 173.9 (C), 51.4 (CH), 40.2 (CH_2), 36.6 (CH_3), 32.5 (2 CH_2), 25.0 (2 CH_2). IR (cm^{-1}) 2922, 2854, 1745, 1457, 1376, 1231, 1203, 1182. HRMS (ESI+) calculated for $C_8H_{15}O_2$ [*M* + H]$^+$ 143.1067, found 143.1064.

2-(Thiophen-2-yl)acetohydrazide (H1):



This compound was synthesized according to **GP-B**, starting with ester **E1** (0.245 g, 1.57 mmol). The crude was purified by flash column chromatography (CH_2Cl_2/MeOH 97:3) to afford hydrazide **H1** as a white solid (0.230 g, 1.47 mmol, 94%). M.p. 95–96 °C. ^{1}H-NMR (400 MHz, CD_3OD) δ 7.26–7.25 (m, 1H), 6.94–6.93 (m, 2H), 3.68 (s, 2H). ^{13}C-NMR (101 MHz, CD_3OD) δ 172.0 (C), 137.7 (C), 127.8 (CH), 127.6 (CH), 125.8 (CH), 35.9 (CH_2). IR (cm^{-1}) 3293, 3105, 2915, 1639, 1534, 1430, 1409, 1351, 1304, 1254, 1037, 959, 850, 693, 483. HRMS (ESI+) calculated for $C_6H_9N_2OS$ [*M* + H]$^+$ 157.0430, found 157.0430.

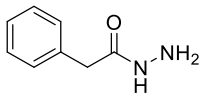
2-(3-Hydroxyphenyl)acetohydrazide (H2):



This compound was synthesized according to **GP-B**, starting with ester **E2** (0.207 g, 1.25 mmol). The crude was purified by flash column chromatography (CH_2Cl_2/MeOH 92:8) to afford hydrazide **H2** as a white solid (0.185 g, 1.11 mmol, 89%). M.p. 171–172 °C. ^{1}H-NMR (400 MHz, DMSO-d_6) δ 9.28 (br s, 1H), 9.16 (br s, 1H), 7.05 (t, *J* = 7.8 Hz, 1H), 6.68–6.64 (m, 2H), 6.61–6.58 (m, 1H), 4.19 (br s, 2H), 3.24 (s, 2H). ^{13}C-NMR (101 MHz, DMSO-d_6) δ 169.5 (C), 157.2 (C), 137.5 (C), 129.0 (CH), 119.6 (CH), 115.9 (CH), 113.3 (CH), 40.5 (CH_2). IR (cm^{-1}) 3311, 3267, 3044, 2915, 1657, 1627, 1588, 1483, 1359, 1293, 1249, 1165, 1025, 887, 757, 713, 687,

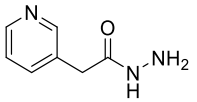
570, 552. HRMS (ESI+) calculated for $C_8H_{11}N_2O_2$ [*M* + H]$^+$ 167.0815, found 167.0815.

2-Phenylacetohydrazide (H3):



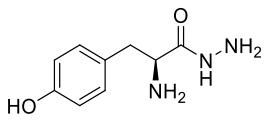
This compound was synthesized according to **GP-B**, starting with methyl phenylacetate (**E3**, 0.150 g, 1.00 mmol). The crude was purified by flash column chromatography (CH_2Cl_2/MeOH 97:3) to afford hydrazide **H3** as a white solid (0.146 g, 0.970 mmol, 97%). M.p. 114–116 °C. ^{1}H-NMR (400 MHz, CD_3OD) δ 7.31–7.27 (m, 4H), 7.26–7.20 (m, 1H), 3.46 (s, 2H). ^{13}C-NMR (101 MHz, CD_3OD) δ 173.1 (C), 136.6 (C), 130.0 (2 CH), 129.5 (2 CH), 127.9 (CH), 41.8 (CH_2). IR (cm^{-1}) 3292, 3198, 3030, 2918, 1643, 1528, 1496, 1350, 1265, 1146, 997, 702, 553, 476. HRMS (ESI+) calculated for $C_8H_{11}N_2O$ [*M* + H]$^+$ 151.0866, found 151.0865.

2-(Pyridin-3-yl)acetohydrazide (H4):



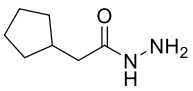
This compound was synthesized according to **GP-B**, starting with ester **E4** (1.04 g, 6.88 mmol). The crude was purified by flash column chromatography (CH_2Cl_2/MeOH from 95:5 to 90:10) to afford hydrazide **H4** as a white solid (0.865 g, 5.72 mmol, 83%). M.p. 122–124 °C. ^{1}H-NMR (400 MHz, CD_3OD) δ 8.48 (d, *J* = 2.2, 1H), 8.43 (dd, *J* = 5.0, 1.7, 1H), 7.80 (dt, *J* = 7.9, 1.7, 1H), 7.42–7.38 (m, 1H), 3.53 (s, 2H). ^{13}C-NMR (101 MHz, CD_3OD) δ 171.9 (C), 150.5 (CH), 148.5 (CH), 138.9 (CH), 133.5 (C), 125.1 (CH), 38.6 (CH_2). IR (cm^{-1}) 3325, 3256, 3215, 3147, 3063, 2904, 1665, 1618, 1580, 1555, 1483, 1428, 1356, 1089, 991, 755, 712, 679, 625, 544. HRMS (ESI+) calculated for $C_7H_{10}N_3O$ [*M* + H]$^+$ 152.0818, found 152.0817.

L-Tyrosine hydrazide (H5):



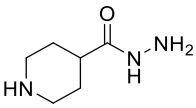
This compound was synthesized according to **GP-B**, starting with L-tyrosine methyl ester (**E5**, 0.195 g, 1.00 mmol). The crude was purified by recrystallization (EtOH) to afford hydrazide **H5** as a white solid (0.160 g, 0.819 mmol, 82%). M.p. 196–198 °C. $[\alpha]_D^{20}$= +47 (*c* = 0.77, MeOH). ^{1}H-NMR (400 MHz, DMSO-d_6) δ 9.13 (s, 1H), 8.86 (br s, 1H), 6.96 (d, *J* = 8.2, 2H), 6.64 (d, *J* = 8.2, 2H), 4.13 (br s, 2H), 3.24 (dd, *J* = 7.8, 5.6, 1H), 2.74 (dd, *J* = 13.3, 5.6, 1H), 2.49–2.45 (m, 1H, overlaps with peak of DMSO), 1.52 (br s, 2H). ^{13}C-NMR (101 MHz, DMSO-d_6) δ 173.7 (C), 155.6 (C), 130.1 (2 CH), 128.7 (C), 114.9 (2 CH), 55.5 (CH), 40.7 (CH_2). IR (cm^{-1}) 3368, 3309, 3290, 3021, 2920, 1644, 1594, 1514, 1455, 1385, 1243, 1191, 1094, 1032, 981, 826, 791, 655, 530. HRMS (ESI+) calculated for $C_9H_{14}N_3O_2$ [*M* + H]$^+$ 196.1081, found 196.1079.

2-Cyclopentylacetohydrazide (H7):



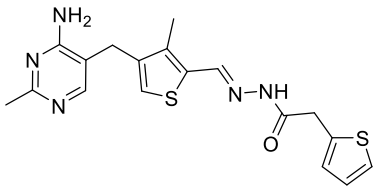
This compound was synthesized according to **GP-B**, starting with ester **E7** (0.100 g, 0.703 mmol). The crude was purified by flash column chromatography (CH_2Cl_2/MeOH 95:5) to afford hydrazide **H7** as a white solid (0.086 g, 0.604 mmol, 86%). M.p. 90–91 °C. ^{1}H-NMR (400 MHz, CD_3OD) δ 2.29–2.18 (m, 1H), 2.15–2.13 (m, 2H), 1.83–1.75 (m, 2H), 1.69–1.52 (m, 4H), 1.22–1.14 (m, 2H). ^{13}C-NMR (101 MHz, CD_3OD) δ 174.9 (C), 41.0 (CH_2), 38.4 (CH), 33.3 (2 CH_2), 25.8 (2 CH_2). IR (cm^{-1}) 3296, 3194, 2951, 2866, 1635, 1534, 1449, 1374, 1218, 1164, 1003, 704, 630, 537, 502. HRMS (ESI+) calculated for $C_7H_{15}N_2O$ [*M* + H]$^+$ 143.1179, found 143.1177.

Piperidine-4-carbohydrazide (H8):



This compound was synthesized according to **GP-B**, starting with methyl piperidine-4-carboxylate (**E8**, 0.143 g, 1.00 mmol). The crude was purified by recrystallization (CH_2Cl_2) to afford hydrazide **H8** as a white solid (35 mg, 0.244 mmol, 24%). M.p. 225–227 °C. ^{1}H-NMR (400 MHz, DMSO-d_6) δ 8.88 (br s, 1H), 4.11 (br s, 2H), 2.92 (dt, *J* = 12.1, 3.4, 2H), 2.41 (td, *J* = 12.1, 3.4, 2H), 2.15–2.07 (m, 1H), 1.52–1.39 (m, 5H). ^{13}C-NMR (101 MHz, DMSO-d_6) δ 174.1 (C), 45.4 (2 CH_2), 40.9 (CH), 29.2 (2 CH_2). IR (cm^{-1}) 3262, 2919, 1645, 1615, 1456, 1340, 1232, 1098, 1009, 927, 740, 587, 462. HRMS (ESI+) calculated for $C_6H_{14}N_3O$ [*M* + H]$^+$ 144.1131, found 144.1129.

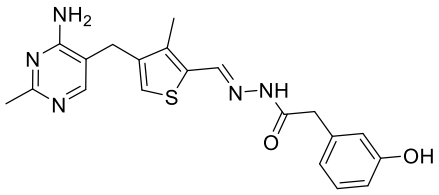
***N'*-((4-((4-Amino-2-methylpyrimidin-5-yl)methyl)-3-methylthiophen-2-yl)methylene)-2-(thiophen-2-yl)acetohydrazide (AH1):**



This compound was synthesized according to **GP-C**, starting with aldehyde **A** (15.0 mg, 0.0606 mmol) and hydrazide **H1** (11.4 mg, 0.0728 mmol). The crude was purified by flash column chromatography (CH_2Cl_2/ MeOH 98:2 to 92:8) to afford **AH1** as a mixture of *E* and *Z* isomers (*E*:*Z* = 43:57) as a white solid (15.4 mg, 0.040 mmol, 66%). ^{1}H-NMR (400 MHz, DMSO-d_6) δ 11.50 (br s, 1H, *E*), 11.25 (br s, 1H, *Z*), 8.47 (s, 1H, *E*), 8.26 (s, 1H, *Z*), 7.65 (s, 1H, *Z*), 7.64 (s, 1H, *E*), 7.39 (dd, 1H, *J* = 4.4, 2.0, *E*), 7.35 (dd, 1H, *J* = 5.0, 1.5, *Z*), 7.08 (s, 1H, *E*), 7.06 (s, 1H, *Z*), 6.98–6.93 (m, 4H, 2*E* and 2*Z*), 6.62 (br s, 4H, 2*E* and 2*Z*), 4.07 (s, 2H, *Z*), 3.73 (s, 2H, *E*), 3.59 (s, 4H, 2*E* and 2*Z*), 2.29 (2s overlap, 6H, 3*E* and 3*Z*), 2.16 (s, 3H, *E*), 2.15 (s, 3H, *Z*). ^{13}C-NMR (101 MHz, DMSO-d_6) δ 170.5 (C), 165.2 (C), 164.76 (C), 164.75 (C), 161.71 (C), 161.70 (C), 154.3 (2 CH), 141.3 (CH), 139.5 (C), 139.4 (C), 139.3 (C), 138.6 (C), 137.7 (CH), 136.7 (C), 136.6 (C), 133.12 (C), 133.08 (C), 126.7 (CH), 126.6 (CH), 126.5 (CH), 126.4 (CH), 125.11 (CH), 125.09 (CH), 124.2 (CH), 123.7

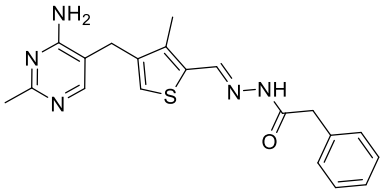
(CH), 111.1 (2 C), 35.6 (CH_2), 33.5 (CH_2), 26.90 (CH_2), 26.89 (CH_2) 25.1 (2 CH_3), 12.25 (CH_3), 12.21 (CH_3). IR (cm^{-1}) 3468, 3304, 3221, 3084, 2925, 1662, 1595, 1556, 1423, 1398, 1232, 1169, 976, 775, 701, 597. HRMS (ESI+) calculated for $C_{18}H_{20}N_5OS_2$ [*M* + H]$^+$ 386.1104, found 386.1074.

***N'*-((4-((4-Amino-2-methylpyrimidin-5-yl)methyl)-3-methylthiophen-2-yl)methylene)-2-(3-hydroxyphenyl)acetohydrazide (AH2):**



This compound was synthesized according to **GP-C**, starting with aldehyde **A** (15.0 mg, 0.0606 mmol) and hydrazide **H2** (12.1 mg, 0.0728 mmol). The crude was purified by flash column chromatography (CH_2Cl_2/MeOH 95:5 to 80:20) to afford **AH2** as a mixture of *E* and *Z* isomers (*E*:*Z* = 47:53) as a white solid (16.0 mg, 0.0405 mmol, 67%). ^{1}H-NMR (400 MHz, DMSO-d_6) δ 11.46 (br s, 1H, *E*), 11.17 (br s, 1H, *Z*), 9.35 (br s, 2H, 1*E* and 1*Z*), 8.50 (s, 1H, *E*), 8.24 (s, 1H, *Z*), 7.64 (s, 2H, 1*E* and 1*Z*), 7.11–7.03 (m, 4H, 2*E* and 2*Z*), 6.72–6.70 (m, 4H, 2*E* and 2*Z*), 6.64–6.58 (m, 6H, 3*E* and 3*Z*), 3.77 (s, 2H, *Z*), 3.58 (s, 4H, 2*E* and 2*Z*), 3.39 (s, 2H, *E*), 2.29 (2s overlap, 6H, 3*E* and 3*Z*), 2.16 (s, 3H, *E*), 2.14 (s, 3H, *Z*). ^{13}C-NMR (101 MHz, DMSO-d_6) δ 171.7 (C), 166.2 (C), 164.7 (2 C), 161.7 (2 C), 157.3 (C), 157.2 (C), 154.3 (2 CH), 141.0 (CH), 139.5 (C), 139.4 (C), 139.1 (C), 138.4 (C), 137.3 (CH), 136.9 (C), 136.8 (C), 133.3 (C), 133.2 (C), 129.2 (CH), 129.1 (CH), 124.0 (CH), 123.5 (CH), 120.0 (CH), 119.6 (CH), 116.2 (CH), 115.9 (CH), 113.6 (CH), 113.4 (CH), 111.1 (2 C), 41.4 (CH_2), 38.8 (CH_2), 26.91 (CH_2), 26.90 (CH_2), 25.1 (2 CH_3), 12.2 (CH_3), 12.2 (CH_3). IR (cm^{-1}) 3449, 3340, 3195, 3063, 2924, 1642, 1591, 1558, 1456, 1353, 1286, 1233, 984, 771, 690, 516. HRMS (ESI+) calculated for $C_{20}H_{22}N_5O_2S$ [*M* + H]$^+$ 396.1489, found 396.1460.

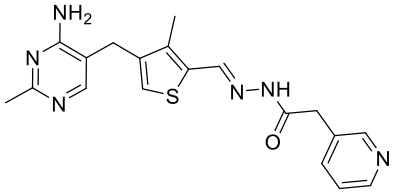
***N'*-((4-((4-Amino-2-methylpyrimidin-5-yl)methyl)-3-methylthiophen-2-yl)methylene)-2-phenylacetohydrazide (AH3):**



This compound was synthesized according to **GP-C**, starting with aldehyde **A** (20.0 mg, 0.081 mmol) and hydrazide **H3** (14.6 mg, 0.097 mmol). The crude was purified by flash column chromatography (CH_2Cl_2/MeOH 95:5 to 90:10) to afford **AH3** as a mixture of *E* and *Z* isomers (*E*:*Z* = 44:56) as a white solid (23.0 mg, 0.0606 mmol, 75%). ^{1}H-NMR (400 MHz, DMSO-d_6) δ 11.46 (br s, 1H, *E*), 11.18 (br s, 1H, *Z*), 8.49 (s, 1H, *E*), 8.24 (s, 1H, *Z*), 7.64 (2s overlap, 2H, 1*E* and 1*Z*), 7.34–7.27 (m, 8H, 4*E* and 4*Z*), 7.26–7.19 (m, 2H, 1*E* and 1*Z*), 7.07 (s, 1H, *E*), 7.05 (s, 1H, *Z*), 6.62 (br s, 4H, 2*E* and 2*Z*), 3.87 (s, 2H, *Z*), 3.58 (s, 4H, 2*E* and 2*Z*), 3.50 (s, 2H, *E*), 2.29 (2s

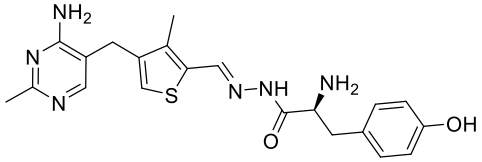
overlap, 6H, 3*E* and 3*Z*), 2.16 (s, 3H, *E*), 2.14 (s, 3H, *Z*). ^{13}C-NMR (101 MHz, DMSO-d_6) δ 171.7 (C), 166.2 (C), 164.76 (C), 164.75 (C), 161.72 (C), 161.71 (C), 154.3 (2 CH), 141.1 (CH), 139.5 (C), 139.4 (C), 139.1 (C), 138.5 (C), 137.4 (CH), 135.64 (C), 135.59 (C), 133.3 (C), 133.2 (C), 129.4 (2 CH), 129.0 (2 CH), 128.3 (2 CH), 128.2 (2 CH), 126.6 (CH), 126.4 (CH), 124.0 (CH), 123.5 (CH), 111.1 (2 C), 41.3 (CH_2), 39.5 (CH_2 overlap with peak of DMSO) ,26.9 (2 CH_2), 25.1 (2 CH_3), 12.23 (CH_3), 12.20 (CH_3). IR (cm^{-1}) 3304, 3214, 3082, 2928, 1661, 1595, 1564, 1477, 1423, 1363, 1172, 1007, 975, 734, 698, 596, 466. HRMS (ESI+) calculated for $C_{20}H_{22}N_5OS$ [*M* + H]$^+$ 380.1540, found 380.1510.

***N'*-((4-((4-Amino-2-methylpyrimidin-5-yl)methyl)-3-methylthiophen-2-yl)methylene)-2-(pyridin-3-yl)acetohydrazide (AH4):**



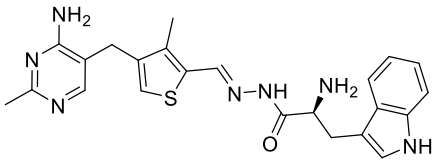
This compound was synthesized according to **GP-C**, starting with aldehyde **A** (20.0 mg, 0.081 mmol) and hydrazide **H4** (14.7 mg, 0.097 mmol). One day later, a second portion of **A** (5.0 mg, 0.020 mmol) was added. Given that the R_f of the product on TLC was the same as the R_f of hydrazide **H4**, full conversion of the hydrazide had to be achieved to enable purification by flash column chromatography. The crude was purified by flash column chromatography (CH_2Cl_2/MeOH 95:5 to 90:10) to afford **AH4** as a mixture of *E* and *Z* isomers (*E*:*Z* = 41:59) as a white solid (32.0 mg, 0.084 mmol, 86%). ^{1}H-NMR (400 MHz, DMSO-d_6) δ 11.55 (br s, 1H, *E*), 11.28 (br s, 1H, *Z*), 8.50–8.42 (m, 5H, 3*E* and 2*Z*), 8.26 (s, 1H, *Z*), 7.73–7.67 (m, 2H, 1*E* and 1*Z*), 7.64 (2s overlap, 2H, 1*E* and 1*Z*), 7.37–7.31 (m, 2H, 1*E* and 1*Z*), 7.07 (s, 1H, *E*), 7.06 (s, 1H, *Z*), 6.62 (br s, 4H, 2*E* and 2*Z*), 3.92 (s, 2H, *Z*), 3.58 (s, 4H, 2*E* and 2*Z*), 3.56 (s, 2H, *E*), 2.29 (2s overlap, 6H, 3*E* and 3*Z*), 2.16 (s, 3H, *E*), 2.15 (s, 3H, *Z*). ^{13}C-NMR (101 MHz, DMSO-d_6) δ 171.1 (C), 165.7 (C), 164.76 (C), 164.75 (C), 161.71 (C), 161.70 (C), 154.3 (2 CH), 150.4 (CH), 150.1 (CH), 147.9 (CH), 147.6 (CH), 141.3 (CH), 139.5 (C), 139.4 (C), 139.2 (C), 138.7 (C), 137.8 (CH), 137.0 (CH), 136.6 (CH), 133.12 (C), 133.11 (C), 131.34 (C), 131.32 (C), 124.1 (CH), 123.6 (CH), 123.4 (CH), 123.3 (CH), 111.1 (2 C), 38.2 (CH_2), 36.3 (CH_2), 26.90 (CH_2), 26.89 (CH_2), 25.1 (2 CH_3), 12.25 (CH_3), 12.21 (CH_3). IR (cm^{-1}) 3320, 3149, 3063, 2928, 1655, 1591, 1574, 1478, 1427, 1361, 1272, 1184, 1102, 973, 783, 709, 630. HRMS (ESI+) calculated for $C_{19}H_{21}N_6OS$ [*M* + H]$^+$ 381.1492, found 381.1462.

(*S*)-2-Amino-*N*'-((4-((4-amino-2-methylpyrimidin-5-yl)methyl)-3-methylthiophen-2-yl)methylene)-3-(4-hydroxyphenyl)propane-hydrazide (AH5):



This compound was synthesized according to **GP-C**, starting with aldehyde **A** (28.7 mg, 0.116 mmol) and hydrazide **H5** (27.1 mg, 0.139 mmol). The crude was purified by flash column chromatography (CH_2Cl_2/MeOH 90:10 to 85:15) to afford **AH5** as a mixture of *E* and *Z* isomers (*E*:*Z* = 57:43) as a white solid (18.0 mg, 0.0424 mmol, 37%). $[\alpha]_D^{20}$= –3.7 (*c* = 0.7, MeOH). ^{1}H-NMR (400 MHz, CD_3OD) δ 8.34 (s, 1H, *E*), 8.13 (s, 1H, *Z*), 7.62 (s, 1H, *Z*), 7.60 (s, 1H, *E*), 7.10–7.04 (m, 6H, 3*E* and 3*Z*), 6.71 (d, *J* = 8.2, 2H, *E*), 6.67 (d, *J* = 8.2, 2H, *Z*), 4.48 (t, *J* = 6.4, 1H, *Z*), 3.68 (s, 2H, *Z*), 3.66 (s, 2H, *E*), 3.52 (t, *J* = 6.9, 1H, *E*), 3.02–2.92 (m, 2H, 1*E* and 1*Z*), 2.84–2.78 (m, 2H, 1*E* and 1*Z*), 2.41 (s, 3H, *Z*), 2.40 (s, 3H, *E*), 2.20 (s, 3H, *Z*), 2.19 (s, 3H, *E*). ^{13}C-NMR (101 MHz, CD_3OD) δ 176.9 (C), 173.0 (C), 166.5 (2 C), 163.68 (C), 163.66 (C), 157.4 (C), 157.3 (C), 154.4 (2 CH), 144.2 (CH), 141.3 (C), 140.3 (C), 140.0 (CH), 139.6 (C), 139.5 (C), 135.2 (C), 134.7 (C), 131.53 (2 CH), 131.45 (2 CH), 129.3 (C), 129.1 (C), 126.5 (CH), 125.5 (CH), 116.3 (2 CH), 116.2 (2 CH), 113.7 (C), 113.6 (C), 57.3 (CH), 53.9 (CH), 41.8 (CH_2), 41.2 (CH_2), 28.29 (CH_2), 28.25 (CH_2), 24.8 (CH_3), 24.7 (CH_3), 12.47 (CH_3), 12.45 (CH_3). IR (cm^{-1}) 3455, 3320, 3201, 3021, 2925, 1658, 1593, 1559, 1514, 1441, 1242, 1173, 1109, 1022, 975, 935, 822, 774, 745, 528. HRMS (ESI+) calculated for $C_{21}H_{25}N_6O_2S$ [*M* + H]$^+$ 425.1754, found 425.1719.

(*S*)-2-Amino-*N*'-((4-((4-amino-2-methylpyrimidin-5-yl)methyl)-3-methylthiophen-2-yl)methylene)-3-(1*H*-indol-3-yl)propanehydrazide (AH6):



This compound was synthesized according to **GP-C**, starting with aldehyde **A** (21.0 mg, 0.085 mmol) and hydrazide **H6** (22.2 mg, 0.102 mmol). The crude was purified by flash column chromatography (CH_2Cl_2/MeOH 93:7 to 85:15) to afford **AH6** as a mixture of *E* and *Z* isomers (*E*:*Z* = 58:42) as a pale yellow solid (12.0 mg, 0.027 mmol, 32%). $[\alpha]_D^{20}$= +11 (*c* = 0.47, MeOH). ^{1}H-NMR (400 MHz, CD_3OD) δ 8.25 (s, 1H, *E*), 8.20 (s, 1H, *Z*), 7.74 (d, *J* = 7.8 , 1H, *Z*), 7.63–7.59 (m, 3H, 2*E* and 1*Z*), 7.34–7.30 (m, 2H, 1*E* and 1*Z*), 7.15 (s, 1H, *Z*), 7.11 (s, 1H, *E*), 7.09–7.05 (m, 4H, 2*E* and 2*Z*), 7.02–6.98 (m, 2H, 1*E* and 1*Z*), 4.63 (m, 1H, *Z*), 3.69–3.66 (m, 5H, 3*E* and 2*Z*), 3.35–3.21 (m, overlaps with peak of CD_3OD, 2H, 1*E* and 1*Z*), 3.09 (dd, *J* = 14.1, 6.3, 1H, *E*), 3.02–2.94 (m, 1H, *Z*), 2.41 (s, 3H, *Z*), 2.40 (s, 3H, *E*), 2.21 (s, 3H, *Z*), 2.15 (s, 3H, *E*). ^{13}C-NMR (101 MHz, CD_3OD) δ 177.2 (C), 173.5 (C), 166.50 (C),

166.48 (C), 163.69 (C), 163.67 (C), 154.44 (CH), 154.41 (CH), 144.3 (CH), 141.2 (C), 140.4 (CH and C), 139.6 (C), 139.5 (C), 138.2 (C), 138.1 (C), 135.0 (C), 134.8 (C), 128.8 (C), 128.7 (C), 126.4 (CH), 125.6 (CH), 124.92 (CH), 124.85 (CH), 122.5 (2 CH), 119.9 (CH), 119.79 (CH), 119.77 (CH), 119.48 (CH), 113.7 (C), 113.6 (C), 112.3 (CH), 112.2 (CH), 111.1 (C), 111.0 (C), 56.4 (CH), 52.6 (CH), 32.4 (CH_2), 32.0 (CH_2), 28.3 (CH_2), 28.2 (CH_2), 24.76 (CH_3), 24.74 (CH_3), 12.47 (CH_3), 12.45 (CH_3). IR (cm^{-1}) 3311, 3188, 3108, 3057, 2923, 1658, 1626, 1592, 1558, 1456, 1434, 1341, 1233, 1099, 972, 926, 808, 741, 588, 423. HRMS (ESI+) calculated for $C_{23}H_{26}N_7OS$ $[M + H]^+$ 448.1914, found 448.1876.

5.4.3 Expression and purification of ThiT

The expression and purification of wild-type, substrate-free ThiT were performed as described previously (Chapter 2, Section 2.4.3).[32]

5.4.4 DCC experiments

Buffers were prepared as follows:
- buffer pH 5.0: McIlvaine's system (citric acid (0.1 M) and Na_2HPO_4 (0.2 M)).
- buffer pH 7.0: potassium phosphate buffer (KP_i (pH 7.0, 50 mM), KCl (150 mM), *n*-decyl-β-D-maltopyranoside (DM, Anatrace, 0.15%, w/v)), prepared from stock solutions of KP_i (pH 7.0, 1.0 M, using K_2HPO_4 (1.0 M) and KH_2PO_4 (1.0 M)), KCl (2.0 M) and DM (20%, w/v).
Deuterated buffers were prepared in the same way but using D_2O instead of H_2O, and adjusting the pH to pH 4.6 (pD 5.0) and pH 6.6 (pD 7.0).

5.4.4.1 Affinity chromatography: nickel-sepharose column

Control experiments without ThiT:

A. Preparation of the DCL. First, the library of 8 acylhydrazones was prepared by mixing buffer pH 5.0 (950 µL), DMSO (26 µL), 8 hydrazides **H1–H8** (8 x 2 µL each, stock solutions of 100 mM in DMSO) and aldehyde **A** (8 µL, stock solution of 100 mM in DMSO). The mixture was incubated at room temperature for 24 h using a rotary mixer.

B. Nickel-sepharose column. The DCL (0.5 mL) and buffer pH 7.0 (7 mL) were added to the column (nickel-sepharose resin, column volume = 0.5 mL), and it was incubated at 4 °C for 30 min with gentle rocking. Then, the column was left to settle down and the flow-through (*ca.* 7 mL) was collected. The elution of the column was done in two different ways:

- Control experiment with two buffers of different concentration of imidazole: first, wash buffer (buffer pH 7.0 containing 50 mM imidazole, 10 mL) was added, and 3 fractions were collected. Then, the elution buffer (buffer pH 7.0 containing 500 mM imidazole, 1.6 mL) was added, and 6 fractions were collected.

- Control experiment eluting with an imidazole gradient: first, the buffer with a lower concentration of imidazole (buffer pH 7.0 containing 50 mM imidazole, 2 mL) was added, and 3 fractions were collected. Then the next buffer (pH 7.0 containing 100 mM imidazole, 2 mL) was added, and 3 fractions were collected. The same was done with the remaining eight buffers of increasing concentration of imidazole (until 500 mM, 2 mL of each buffer, 3 factions collected for each buffer).

C. UPLC-TOF analysis. UPLC-TOF was performed using a Waters Acquity UPLC H-class system coupled to a Waters Xevo-G2 TOF; and a column Acquity UPLC BEH Shield RP18, 1.7 µm, 2.1 x 150 mm. Positive-ion mass spectra were acquired using electrospray ionization (ESI+). Method: run time 20 min, λ = 315 nm, flow rate of 0.3 mL min^{-1}, H_2O/CH_3CN (0.1 % TFA) from 95:5 to 60:40 in 12 min, to 5:95 in 1 min, hold for 1 min, to 95:5 in 1 min, hold for 5 min.

Control experiment with ThiT and without DCL:

ThiT was first allowed to bind to the nickel-sepharose column by incubation of ThiT (200 µL of ~10 µM) with buffer pH 7.0 (9 mL) for 20 min. Then, it was eluted first with one fraction of wash buffer (pH 7.0 containing 250 mM imidazole, 700 µL) followed by one fraction of elution buffer (pH 7.0 containing 500 mM imidazole, 700 µL). The protein concentration in the first (250 mM) elution fraction was 0.023 mg mL^{-1} and in the second (500 mM) elution fraction 0.070 mg mL^{-1}.

5.4.4.2 Size-exclusion chromatography (SEC)

A. Preparation of the DCL. The DCL was prepared in the same way as for the nickel-sepharose column (see Section 5.4.4.1). Before the next step, the pH was set to pH 7.0 by addition of an aqueous solution of KOH (4.0 M).

B1. Control experiment without ThiT. The DCL solution (0.5 mL) was added to buffer pH 7.0 (0.25 mL), and it was incubated at 4 °C for 2 h with gentle rocking. The mixture was centrifuged for 5 min.

B2. Incubation of the DCL with ThiT. The DCL solution (0.5 mL) was added to a solution of ThiT (pH 7.0, 10.7 μM, 0.25 mL), so the final concentration of ThiT was ~3.6 μM, and it was incubated at 4 °C for 2 h with gentle rocking. The mixture was centrifuged for 5 min, and aggregates appeared in the bottom of the Eppendorf tube, which were discarded.

C. SEC. SEC was performed using a column Superdex 200 10/300 GL (GE Healthcare), buffer pH 7.0 as eluent and measuring the absorbance at λ = 280 nm. Fractions of 0.5 mL were collected.

D. UPLC-TQD-SIR analysis. UPLC-TQD-SIR was performed using a Waters Acquity UPLC system coupled to a Waters TQD; and a column Acquity UPLC HSS T3, 100 Å, 1.8 μm, 2.1 mm x 150 mm. Positive-ion mass spectra were acquired using electrospray ionization (ESI+). Method: run time 20 min, λ = 315 nm, flow rate of 0.3 mL min^{-1}, H_2O/CH_3CN (0.1% formic acid) from 95:5 to 60:40 in 13 min, to 5:95 in 1 min, hold for 1 min, to 95:5 in 1 min, hold for 4 min.

5.4.4.3 ^{1}H-STD-NMR spectroscopy

General remarks. ^{1}H-STD-NMR experiments were performed on a Varian Inova 600 MHz spectrometer equipped with a 5 mm indirect detection probe head, at a temperature of 25 °C. Selective saturation was achieved by a train of Gauss-shaped pulses of 50 ms each, separated by a 0.1 ms delay. A number of 60 selective pulses were applied, leading to a total length of the saturation train of 3 s. The *on*-resonance irradiation on the protein was performed at a chemical shift of –1 or –2 ppm for the control experiments with a known binder, and –1.1 ppm for the experiment with the DCL with protein, and the *off*-resonance irradiation was set to –25 ppm in all cases, where no protein signals were present. The number of scans used was 8192 (4096 for *on*- and 4096 for *off*-resonance). NMR spectra were multiplied by an exponential line broadening function of 1 Hz prior to Fourier transformation. All spectra were recorded with a 20 ms spin-lock pulse, which minimizes the background protein resonances. The 'DPFGSE sculpted solvent suppression' was enabled. The data were acquired interleaved with blocks of 4 scans. The spectra were subtracted manually in MestReNova.

Control experiment with a known binder of ThiT:

First, a buffer-exchange column was used to transfer ThiT from the buffer pH 7.0 (0.5 mL, 12.8 μM) to the same buffer but with D_2O (pD 7.0), using a column illustra NAP-5 (GE Healthcare), which afforded ~400 μL of 9.8 μM ThiT. To this solution of ThiT, compound **B1** (4 μL, stock solution of 20 mM in

DMSO-d_6), buffer pD 7.0 (75 µL) and DMSO-d_6 (21 µL) were added. Given that a 20-fold excess of compound **B1** with respect to the protein was not enough to obtain a good ^{1}H-STD-NMR spectrum, the concentration was increased to 100-fold by addition of **B1** (3.2 µL, stock solution of 100 mM in DMSO-d_6).

Experiment with ThiT and DCL:

First, a buffer-exchange column was used to transfer ThiT from pH 7.0 (0.7 mL, 20.7 µM) to the same buffer but with D_2O (pD 7.0), using a column illustra NAP-5 (GE Healthcare), which afforded ~950 µL of 10.0 µM ThiT.

Knowing the volume and concentration of ThiT available for the experiment, the amount of building blocks was calculated to have 100-fold excess of aldehyde **A**. In each DCL, 1.2 eq of each hydrazide with respect to the aldehyde were added (there were 4 hydrazides in each DCL: in total 4.8 eq of hydrazides with respect to the aldehyde).

For each DCL, 450 µL of 10.0 µM ThiT were used, in a final volume of 550 µL, which gives a final concentration of 8.2 µM ThiT in the NMR tube. Therefore, the final concentration of aldehyde (100-fold) is 820 µM and of each hydrazide is 984 µM. To achieve these final concentrations, each DCL of 100 µL has 4.51 mM aldehyde **A** and 5.42 mM of each hydrazide **H1–H8**. Considering also that the maximum concentration of DMSO tolerated by ThiT is 5% (27.5 µL in 550 µL), each DCL was prepared as follows: buffer pD 5.0 (72.5 µL), aldehyde **A** (2.3 µL, stock solution of 200 mM in DMSO-d_6), 4 hydrazides (4 x 2.7 µL each, stock solutions of 200 mM in DMSO-d_6) and DMSO-d_6 (14.4 µL). For DCL-A, hydrazides **H1**, **H2**, **H6** and **H8** were included, and for DCL-B, hydrazides **H3**, **H4**, **H5** and **H7**. Each DCL was incubated at room temperature for 24 h in a rotary mixer, and then added to ThiT. Subsequently, the ^{1}H-STD-NMR spectrum were recorded.

For each DCL, a control experiment was carried out: an identical DCL was added to 450 µL of buffer pD 7.0 (without ThiT), and the ^{1}H-STD-NMR spectrum was recorded to check that there were no signals in the absence of protein.

To identify which acylhydrazones appeared in the ^{1}H-STD-NMR spectrum, a sample of each acylhydrazone was prepared individually: hydrazide (2 µL, stock solution of 200 mM in DMSO-d_6), aldehyde **A** (10 µL, stock solution of 200 mM in DMSO-d_6), DMSO-d_6 (13 µL) and buffer pD 5.0 (75 µL). After incubation at room temperature for 24 h in a rotary mixer, buffer pD 7.0 (450 µL) was added, and the ^{1}H-NMR spectra were recorded on a Varian AMX 400 spectrometer at 25 °C.

5.4.5 Binding-affinity determination

The binding affinities of the acylhydrazones **AH1**, **AH3**, **AH4** and **AH5** for ThiT were determined by ITC as described previously (Chapter 3, Section 3.4.4).

5.5 Contributions from co-authors

The expression and purification of protein, as well as the binding affinity measurements were performed by L. J. Y. M. Swier from the group of Prof. D. J. Slotboom.

5.6 References

(1) Huc, I.; Lehn, J.-M. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 2106.

(2) Ramström, O.; Lehn, J.-M. *Nat. Rev. Drug Discov.* **2002**, *1*, 26.

(3) Ladame, S. *Org. Biomol. Chem.* **2008**, *6*, 219.

(4) Herrmann, A. *Chem. Soc. Rev.* **2014**, *43*, 1899.

(5) Mondal, M.; Hirsch, A. K. H. *Chem. Soc. Rev.* **2015**, *44*, 2455.

(6) Ulrich, S.; Dumy, P. *Chem. Commun.* **2014**, *50*, 5810.

(7) Eugene H Cordes; William P Jencks. *J. Am. Chem. Soc.* **1962**, *84*, 826.

(8) Bhat, V. T.; Caniard, A. M.; Luksch, T.; Brenk, R.; Campopiano, D. J.; Greaney, M. F. *Nat. Chem.* **2010**, *2*, 490.

(9) van der Vlag, R.; Hirsch, A. K. H. Analytical methods in protein-templated dynamic combinatorial chemistry, In: Reference Module in Chemistry, Molecular Sciences and Chemical Engineering 2016, from *Comprehensive Supramolecular Chemistry II*, Oxford: Elsevier (in press).

(10) Fang, Z.; He, W.; Li, X.; Li, Z.; Chen, B.; Ouyang, P.; Guo, K. *Bioorg. Med. Chem. Lett.* **2013**, *23*, 5174.

(11) He, W.; Fang, Z.; Yang, Z.; Ji, D.; Chen, K.; Guo, K. *RSC Adv.* **2015**, *5*, 23224.

(12) Yang, Z.; Fang, Z.; He, W.; Wang, Z.; Gan, H.; Tian, Q.; Guo, K. *Bioorg. Med. Chem. Lett.* **2016**, *26*, 1671.

(13) Ramström, O.; Lehn, J.-M. *ChemBioChem* **2000**, *1*, 41.

(14) McNaughton, B. R.; Miller, B. L. *Org. Lett.* **2006**, *8*, 1803.

(15) Besenius, P.; Cormack, P. A. G.; Ludlow, R. F.; Otto, S.; Sherrington, D. C. *Org. Biomol. Chem.* **2010**, *8*, 2414.

(16) Gasparini, G.; Rastrelli, F.; Prins, L. J. *Org. Biomol. Chem.* **2013**, *11*, 6580.

(17) Gareiss, P. C.; Sobczak, K.; McNaughton, B. R.; Palde, P. B.; Thornton, C. A.; Miller, B. L. *J. Am. Chem. Soc.* **2008**, *130*, 16254.

(18) Hochgürtel, M.; Kroth, H.; Piecha, D.; Hofmann, M. W.; Nicolau, C.; Krause, S.; Schaaf, O.; Sonnenmoser, G.; Eliseev, A. V. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 3382.

(19) Hochgürtel, M.; Biesinger, R.; Kroth, H.; Piecha, D.; Formann, M. W.; Krause, S.; Schaaf, O.; Nicolau, C.; Eliseev, A. V. *J. Med. Chem.* **2003**, *46*, 356.

(20) Caraballo, R.; Dong, H.; Ribeiro, J. P.; Jiménez-Barbero, J.; Ramström, O. *Angew. Chem. Int. Ed.* **2010**, *49*, 589.

(21) Sindelar, M.; Wanner, K. T. *ChemMedChem* **2012**, *7*, 1678.

(22) Gerber, P. R.; Müller, K. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 251.

(23) LeadIT (version 2.1.8), BioSolveIT GmbH, An Der Ziegelei 79, 53757 Sankt Augustin, Germany, 2014.

(24) Reulecke, I.; Lange, G.; Albrecht, J.; Klein, R.; Rarey, M. *ChemMedChem* **2008**, *3*, 885.

(25) Schneider, N.; Hindle, S.; Lange, G.; Klein, R.; Albrecht, J.; Briem, H.; Beyer, K.; Claußen, H.; Gastreich, M.; Lemmen, C.; Rarey, M. *J. Comput.-Aided Mol. Des.* **2012**, *26*, 701.

(26) Erkens, G. B.; Berntsson, R. P.-A.; Fulyani, F.; Majsnerowska, M.; Vujičić-Žagar, A.; ter Beek, J.; Poolman, B.; Slotboom, D. J. *Nat. Struct. Mol. Biol.* **2011**, *18*, 755.

(27) Mayer, M.; Meyer, B. *Angew. Chem. Int. Ed.* **1999**, *38*, 1784.

(28) Mondal, M.; Radeva, N.; Köster, H.; Park, A.; Potamitis, C.; Zervou, M.; Klebe, G.; Hirsch, A. K. H. *Angew. Chem. Int. Ed.* **2014**, *53*, 3259.

(29) Meinecke, R.; Meyer, B. *J. Med. Chem.* **2001**, *44*, 3059.

(30) Assadi-Porter, F. M.; Tonelli, M.; Maillet, E.; Hallenga, K.; Benard, O.; Max, M.; Markley, J. L. *J. Am. Chem. Soc.* **2008**, *130*, 7212.

(31) Hirsch, A. K. H.; Buhler, E.; Lehn, J.-M. *J. Am. Chem. Soc.* **2012**, *134*, 4177.

(32) Swier, L. J. Y. M.; Monjas, L.; Guskov, A.; de Voogd, A. R.; Erkens, G. B.; Slotboom, D. J.; Hirsch, A. K. H. *ChemBioChem* **2015**, *16*, 819.